I

ANALYSIS AND METRIC GEOMETRY

LINE INTEGRALS, THEIR SEMICONTINUITY PROPERTIES, AND THEIR INDEPENDENCE OF THE PATH

1. THE LENGTH (RIEMANN INTEGRATION, LEBESGUE INTEGRATION, WEIERSTRASS-ARCHIMEDES SUMS)

ONE of the oldest and best known line integrals is the length of a curve. According to the classical theory, the curve C in the (t, y)-plane given by the equation $y = y(t)$, $(a \leq t \leq b)$ has the length

$$l(C) = \int_a^b \sqrt{1 + (y'(t))^2}\,dt.$$

According to Riemann's definition of a definite integral, $l(C)$ has the following meaning: Let T be a finite ordered subset of $[a, b]$, say

$$T = \{t_0, t_1, \cdots t_{n-1}, t_n\},$$

where $t_0 = a$, $t_n = b$, and $t_i < t_{i+1}$. We call the greatest of the n numbers $t_{i+1} - t_i$ the norm of T and denote it by $\nu(T)$. Let T^* be a set $\{t_0, t_0^*, t_1, t_1^*, \cdots t_{n-1}, t_{n-1}^*, t_n\}$ where $t_i \leq t_i^* \leq t_{i+1}$. By the *Riemann sum* associated with T^*, we mean the sum

$$R(T^*) = \sum_{i=0}^{n-1} \sqrt{1 + (y'(t_i^*))^2}\,(t_{i+1} - t_i).$$

The length $l(C)$ is that number (finite or $+\infty$) to which the numbers $R(T^*)$ come as close as we please for all T^* for which $\nu(T)$ is sufficiently small—provided that a number of this kind exists. It does exist for the curve $y=y(t)$ if and only if the integrand $\sqrt{1+(y'(t))^2}$ or, what is equivalent, the function $y'(t)$ is continuous almost everywhere (i.e., for all t outside of a set of measure 0 in Lebesgue's sense).

In particular, the integrand and hence the derivative $y'(t)$ must exist everywhere in $[a, b]$; that is to say, the curve C must have a tangent at each point. By a well-known procedure of splitting the domain of integration into parts, we may admit a finite set of numbers t for which $y'(t)$ is not defined, and even certain simple infinite sets of points at which the curve C has no tangents.

At any rate, the classical theory ties up the definition of the length of a curve with the question of whether the curve has tangents and how these tangents vary.

This is unsatisfactory, because the primitive geometric ways of determining the length of a curve do not make use of tangents to the curve. Our intuitive idea of the length has no connection with that of tangents. Moreover, we are interested in the length of any curve regardless of whether or not it has tangents, and we actually can determine the length of curves which do not have tangents. There are two ways of severing the ties between length and tangents.

One attempt consists in interpreting the $\int$ sign in $l(C) = \int_a^b \sqrt{1+(y'(t))^2}dt$ as a Lebesgue integral. The Lebesgue integral of a function in an interval may exist without the function being defined everywhere in the interval. It is sufficient that the integrand is defined almost everywhere provided that, on its domain of definition, it satisfies certain conditions. Now if a curve C given by $y=y(t)$ $(a \leq t \leq b)$ has a finite length $l(C)$ in the sense of geometry, then Lebesgue proved that the set of all t for which C has no tangent is of measure 0. Thus, for such a curve, $y'(t)$ and $\sqrt{1+(y'(t))^2}$ do exist almost everywhere. Moreover, Lebesgue proved that if $y(t)$ defines a curve of finite length then the function $\sqrt{1+(y'(t))^2}$ has a finite Lebesgue integral. But with regard to the relation between this integral and the length $l(C)$, all that can be said is that the integral is never greater than the length. It may be actually smaller. Tonelli proved that the equality $l(C)=\int_a^b\sqrt{1+(y'(t))^2}dt$ holds if and only if the function $y(t)$ is absolutely continuous.[1]

For instance, let C be the curve joining the points $(0, 0)$ and $(1, 1)$ given by the equation $y=y(t)$ $(0\le t\le 1)$, where $y(t)$ is the continuous but not absolutely continuous function, assuming the value $1/2$ for $1/3\le t\le 2/3$, the values $1/4$ for $1/9\le t\le 2/9$ and $3/4$ for $7/9\le t\le 8/9$, etc., in general the value

$$y=\sum_{k=1}^{n-1}\frac{a_k}{2^k}+\frac{1}{2^n}\quad\text{for}\quad\sum_{k=1}^{n}\frac{2a_k}{3^k}+\frac{1}{3^n}\le t\le\sum_{k=1}^{n}\frac{2a_k}{3^k}+\frac{2}{3^n}$$

$$(a_1,\ a_2,\ \cdots\ a_{n-1}=0\text{ or }1).$$

This function $y(t)$ being constant on intervals whose total length is 1 has the derivative 0 almost everywhere in $[0, 1]$. Thus the Lebesgue integral

$$\int_0^1\sqrt{1+(y'(t))^2}dt=\int_0^1 1\cdot dt=1.$$

This number is even smaller than the length $\sqrt{2}$ of the straight segment between the endpoints $(0, 0)$ and $(1, 1)$ of the curve C. The length of C is 2.

Thus, by interpreting $\int$ as a Lebesgue integral, we cannot uphold the definition of the length of a curve as the line integral $\int\sqrt{1+(y'(t))^2}dt$ beyond a special class of curves of finite length, viz., the curves defined by absolutely continuous functions—although the line integral itself can be formed for more general curves. Moreover, this method digresses from the geometric intuition that accompanies our elementary determination of length.

The second way out of the difficulty consists in considering what is called Weierstrass sums instead of Riemann sums. By the *Weierstrass sum* associated with T, we mean an expression similar to a Riemann sum, but involving difference quotients instead of differential quotients, namely

$$W(T)=\sum_{i=0}^{n-1}\sqrt{1+\left(\frac{y(t_{i+1})-y(t_i)}{t_{i+1}-t_i}\right)^2}\,(t_{i+1}-t_i).$$

If for $y(t)$ the mean value theorem of differential calculus holds, then for each T there exists a T^* such that $R(T^*)=W(T)$. Hence, if there is a number $l(C)$ to which $R(T^*)$ converges when $\nu(T)$ converges toward zero, then this number $l(C)$ is also the limit of $W(T)$ when $\nu(T)$ converges toward zero. But there may exist a number $l(C)$ to which $W(T)$ converges with $\nu(T)\to 0$ without $R(T^*)$ approaching a limit. In fact, one can prove for each curve C which is given by a continuous function $y=y(t)$, that there is a number $l(C)$ finite or $+\infty$ to which $W(T)$ converges with $\nu(T)\to 0$. However, as we saw, for many continuous functions $y(t)$, there is no number to which $R(T^*)$ would converge with $\nu(T)\to 0$. E.g., there is no such number $l(C)$ if the continuous function $y(x)$ is nowhere differentiable, for the simple reason that in this case the numbers $R(T^*)$ themselves do not exist.

The second way of severing the definition of the length of a curve from its differentiability properties consists in calling length of C the limit of the numbers $W(T)$ rather than that of the numbers $R(T^*)$.

If we examine $W(T)$, we see that it is equal to

$$\sum_{i=0}^{n-1}\sqrt{(t_{i+1}-t_i)^2+(y(t_{i+1})-y(t_i))^2}.$$

In the Euclidean plane the i-th summand of this sum is the distance between p_i and p_{i+1}, where we denote by p_i the point with the coordinates $(t_i, y(t_i))$. Thus the sum $W(T)$ is the length of the polygon $\{p_0, p_1, \cdots p_n\}$ inscribed in the curve C. It is what may be called *Archimedes Sum*

$$A(T) = \sum_{i=0}^{n-1} \text{dist } (p_i, p_{i+1}),$$

for it is these lengths of inscribed polygons that, in special cases, Archimedes used for the determination of the length of a curve C.

Now when we spoke about primitive geometric methods of determining the length of a curve, we meant the limit to which the lengths $A(T)$ of the inscribed polygons converge when the norm of the polygons tends toward 0. And when the theory of length as a Lebesgue line integral states that $\int_a^b \sqrt{1 + (y'(t))^2}dt$ exists for each curve of finite length, and admits that this integral may be smaller than the *geometric* length of the curve, it is again this limit of the $A(T)$ which is meant. Thus the identity of $W(T)$ with $A(T)$ seems to indicate that the Weierstrass sums are a tool of analysis more adequate to the geometric problem than the Riemann or Lebesgue sums. At the same time, a theory based on the $W(T)$ is the most radical way of severing the concept of length from that of tangents, and of freeing the analysis from unessential differentiability assumptions.

2. THE GENERAL LINE INTEGRAL AND FOUR MAIN PROBLEMS CONCERNING IT

Now we consider the general line integral. In the following we shall restrict ourselves to curves in the Euclidean plane since the generalization to curves of higher dimensional Euclidean spaces does not present any difficulty. In the plane we shall admit any curve C given by a parametric representation $x = x(t)$, $y = y(t)$ $(a \leq t \leq b)$. A line integral is determined by a function of four real variables $F(x,y,u,v)$. The value of the line integral of F along the curve C is

$$J(C) = \int_a^b F(x(t),\, y(t),\, x'(t),\, y'(t))\,dt.$$

Some of the main problems concerning these general line integrals are the following questions:

I. Under what conditions concerning the integrand F and the curve C does $J(C)$ *exist*?

II. Under what conditions concerning F is the integral a *semicontinuous* functional of C, say lower semicontinuous, i.e., a functional with the property that for each curve C and each $\epsilon > 0$ there exists a neighborhood of C such that for each curve C_1 located in this neighborhood we have $J(C_1) > J(C) - \epsilon$?

III. Under what conditions concerning a class of curves and the integrand F does the class contain a curve *minimizing* $J(C)$, i.e., a curve for which the value of J does not surpass the value of J for any other curve of the class?

IV. Under what conditions concerning the integrand F is the integral $J(C)$ *independent of the path*, that is to say, assumes the same value for any two coterminal curves?

3. THE ANSWER TO THE FUNDAMENTAL QUESTIONS ON THE BASIS OF RIEMANN INTEGRATION

The classical treatment of these problems is based on Riemann integration. From this point of view $J(C)$ is the limit of Riemann sums

$$R(T^*) = \sum_{i=0}^{n-1} F(x(t_i^*),\, y(t_i^*),\, x'(t_i^*),\, y'(t_i^*))(t_{i+1} - t_i)$$

when $\nu(T) \to 0$. In order that such a limit exists, it is necessary and sufficient that the function $F(x(t), y(t), x'(t), y'(t))$ is almost everywhere continuous. To guarantee this condition, we must not only know that F has certain continuity properties, but also that $x'(t)$ and $y'(t)$ exist and are continuous to a large extent. In answer to question I, the classical theory did not hesitate to make these assumptions, but, of course, this procedure is no more satisfactory in the general case, than it is in the case of length.

It is clear that similar restrictions would have to be imposed on F and C in order to guarantee that the Riemann integral is a semicontinuous functional and to prove the existence of minimizing curves in answer to questions II and III, although these questions seem not to have been studied extensively for Riemann integrals.

To question IV the classical theory gave two answers. Let $u(x, y)$ and $v(x, y)$ be two continuous functions. Then $\int u dx + v dy$ is independent of the path if either one of the two following conditions is satisfied:

1. The partial derivatives $\frac{\partial u}{\partial y}$ and $\frac{\partial v}{\partial x}$ exist, are continuous, and are equal to each other at each point (x, y) of a closed domain.

2. $\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x}$, and both functions u and v have differentials at each point. A function $g(x, y)$ is said to have a differential at the point (x_0, y_0), if there exist two numbers $a(x_0, y_0)$ and $b(x_0, y_0)$ with the following property: for each $\epsilon > 0$ there is a $\eta > 0$ such that

$$\begin{aligned} g(x, y) &= g(x_0, y_0) + a(x_0, y_0)(x - x_0) \\ &\quad + b(x_0, y_0)(y - y_0) + \zeta(x,y)(|x - x_0| + |y - y_0|) \\ \text{and } &|\zeta(x, y)| < \epsilon \end{aligned}$$

for each (x, y) for which $|x - x_0| + |y - y_0| < \eta$. The numbers $a(x_0, y_0)$ and $b(x_0, y_0)$ are, of course, the partial derivatives of g for (x_0, y_0) with respect to x and y, respectively.

The sufficiency of condition 2 was proved by Heffter, following Goursat's method in the theory of the functions of a complex variable. This condition has a local character since the dependence of η on ϵ may vary with (x_0, y_0). No uniformity of the extent to which $g(x, y)-g(x_0, y_0)$ is approximated with a prescribed degree of accuracy by the linear terms $a(x_0, y_0)(x-x_0)+b(x_0, y_0)(y-y_0)$, is postulated. The older condition 1, by assuming continuity of the partial derivatives in a bounded and closed domain, does imply uniform continuity of these partial derivatives. Exceptional points at which the differentials or the partial derivatives do not exist or the partial derivatives are discontinuous and unequal are admissible on the basis of Riemann integration, provided that the set of all exceptional points has the content 0, i.e., may be covered by a finite number of squares of an arbitrarily small total area.[2]

Summarizing, we can say: with regard to all four problems, the classical theory based on Riemann integration made many assumptions about the differentiability of the integrand F and the curve C which are required by the method rather than by the questions themselves. For, the questions concerning the line integrals can be formulated with regard to integrands that are not differentiable and curves that have no tangents.

4. THE ANSWER TO THE FOUR FUNDAMENTAL QUESTIONS ON THE BASIS OF LEBESGUE INTEGRATION

In the theory of length we saw two ways out of this difficulty. With regard to general line integrals, the first of these ways was followed. One considered Lebesgue line integrals.

Hahn seems to have been the first to introduce Lebesgue integrals into the calculus of variations.[3] Tonelli based his entire book[1] on Lebesgue integration. The fundamental concepts underlying this work which marked one of the greatest steps forward in its field can be summarized as follows:

a. *The Integrand.* Tonelli starts by making Weierstrass's assumption that the integrand $F(x, y, u, v)$ is positively homogeneous with regard to u and v, i.e., satisfies the equality $F(x, y, ku, kv) = kF(x, y, u, v)$ for each $k > 0$ and each quadruple x, y, u, v for which (x, y) lies in a certain domain of the plane. This assumption is made in order to guarantee that for any two different parametric representations of the same curve, the integral assumes the same value. It implies that

$$\frac{F_{uu}}{v^2} = -\frac{F_{uv}}{uv} = \frac{F_{vv}}{u^2}$$

where we write F_{uu} for $\frac{\partial^2 F}{\partial u^2}$, etc. Calling $F_1\,(x, y, u, v)$ the common value of the above three quotients Tonelli assumes that $F_1 \geq 0$ or, as he says, that F is *positively quasi-regular.* Furthermore, he assumes that the function $F(x, y, u, v)$ is *continuous* with respect to the quadruples of variables. Moreover, F is assumed to admit at least *first partial derivatives* with respect to the last two variables. For part of his theory Tonelli assumes the existence of second derivatives, but remarks that other parts are independent of this assumption. For instance, the hypothesis of quasi-regularity which involves second partial derivatives can be replaced by the assumption that for each point (x_0, y_0) the figurative of F is convex downward. The figurative of F for the point (x_0, y_0), introduced by Hadamard, is the surface $z = F(x_0, y_0, u, v)$ in the (u, v, z)-space. On account of the positive homogeneity of F, this surface is a cone with the vertex at the origin. The cone degenerates into a plane if and only if $F_1(x_0, y_0, u, v) = 0$ for all values of u and v.

A subsequent idea due to McShane[4] and Aronszajn[5] allows us to dispense even with continuity of the integrand F provided that F is lower semicontinuous, and the figurative convex at each point.

b. *The Admissible Curves.* The curves along which the Lebesgue line integral $J(C)$ may be formed are the curves of finite length, if this length is used as parameter. That is to say, we can form $\int_0^{l(C)} F(x(s), y(s), x'(s), y'(s))ds$ for each curve C given by two equations $x = x(s)$, $y = y(s)$ $(0 \leq s \leq l(C))$ where s denotes the length of the segment of the curve C between the initial point $(x(0), y(0))$ and the point $(x(s), y(s))$. If C is given by two equations $x = x(t)$, $y = y(t)$ $(a \leq t \leq b)$, then Tonelli requires absolute continuity of the functions $x(t)$, $y(t)$, a condition which is automatically satisfied by the functions $x(s)$, $y(s)$ defining a curve of finite length.

Under the specified conditions concerning the integrand F, Tonelli proves that $J(C)$, the integral of F, is lower semicontinuous for all curves C of uniformly bounded lengths. Under slightly stronger conditions concerning F, he proves that the integral is lower semicontinuous for all curves of finite length. The additional assumption concerning F (called *seminormality* of F) is that for no point (x_0, y_0) is $F_1(x_0, y_0, u, v) = 0$ for all (u, v) or, in other words, that for no point is the figurative of F a plane. Tonelli even admits points (x_0, y_0) for which the figurative is a plane provided that each of these points lies within a neighborhood such that $J(C) \geq 0$ for any closed curve C of finite length contained in the neighborhood.

c. *The Functional of Comparison.* As we saw when speaking about the admissible curves, the integral along a curve is frequently related to the length of the curve. The integral is merely defined for curves of finite length; some semicontinuity properties are merely proved on the sets of curves of uniformly bounded lengths; and, as we shall see, Tonelli's theorem about the existence of curves that minimize the integral, contains another reference to the length. For these reasons we shall call the length of curves the functional of comparison used in Tonelli's theory.

d. *The Space.* All the curves considered in Tonelli's theory are contained in a Euclidean or at least a Cartesian space in which each point can be described by n real coordinates.

In answer to question III Tonelli states conditions sufficient for the existence in each closed class of curves of a curve C_0 that minimizes the integral $J(C)$. From the lower semicontinuity of the integral he derives the existence of such a minimizing curve under the additional hypothesis that for each finite number M the curves C for which $J(C) \leq M$ are of uniformly bounded lengths. Implicitly underlying Tonelli's deduction is a theorem of the general analysis generalizing the classical theorem of R. Baire that a lower semicontinuous function on a compact domain of definition assumes its minimum.

A construction due to Hahn and later generalized by Carathéodory, Tonelli, Graves and McShane throws light on the additional hypothesis that for each M the curves C for which $J(C) \leq M$, are of uniformly bounded lengths. If this condition is not satisfied (i.e., in case there exists a sequence of curves $C_1, C_2, \cdots$ for which all numbers $J(C_n)$ are $\leq M$ while the lengths of $C_1, C_2, \cdots$ converge toward ∞), then Hahn's construction yields a sequence of curves $D_1, D_2, \cdots$ for which all the numbers $J(D_n)$ are ≤ 0 while the lengths of D_1, D_2, $\cdots$ converge toward ∞. Hence it is sufficient to postulate that the curves C for which $J(C) \leq 0$ are of uniformly bounded lengths.

With regard to problem IV, Montel introduced Lebesgue integration and proved that a line integral $\int u dx + v dy$ is independent of the path provided that $\partial u/\partial y$ and $\partial v/\partial x$ exist and are equal almost everywhere.[6] Montel's successors weakened his assumptions further, in particular with regard to functions of a complex variable, the conditions being formulated in terms of the partial derivatives of u and v.

Of a different type is a condition due to Schauder.[7] Besides continuity of the functions $u(x, y)$ and $v(x, y)$ defined in a rectangle $a \leq x \leq b$, $c \leq y \leq d$, it requires that

$$\int_c^y v(x, \eta)d\eta - \int_a^x u(\xi, y)d\xi,$$

a function of the variables x and y, can be represented as the sum of two functions $\varphi(x)$ and $\psi(y)$ each depending upon one of the two variables only. This very elegant analytic condition does not presuppose differentiability of u and v. On the other hand, it does not seem to give us much of a geometric insight into where the classical arguments are redundant. The work of Besikovic, Looman, Menchoff, Saks and others is based on differentiability assumptions.

In terminating this brief summary of how the four main questions were treated for Lebesgue line integrals, it should be emphasized that with Lebesgue integration, a branch of modern geometry was introduced into Analysis, viz., the geometry of measure. What fundamental role this geometry played in the modern analysis one will appreciate by reading, for instance, Saks's book on integration.[1]

At any rate, the progress marked by the introduction of

Lebesgue integration as compared with the old theory based on Riemann integrals was so great that it can hardly be overestimated. Today, in view of its permanent achievements, the theory developed on this basis during the first two decades of this century can be called classical.

5. EXAMPLES OF CASES EXCLUDED BY THE CLASSICAL THEORIES

It seems to me, however, that there are two objections to this method. The first is that although the method is based on a chapter of geometry, viz., the theory of measure, it is not capable of as simple and intuitive geometric interpretation as, e.g., the theory of length based on the study of inscribed polygons. Facts like the difference between geometric length and the length integral of Lebesgue discussed in section 1 of this paper are merely symptoms of a more general deviation of analysis from geometric intuition in this particular field. But since this argument is one of personal taste rather than one of objective criticism, I shall not insist on it.

The second remark is of a quite objective character. Although by introducing the Lebesgue integral one weakens considerably the assumptions that the theory dealing with Riemann integrals had to make concerning the integrand F and the curve C, one still has to introduce assumptions that have no direct connection with the problem and are introduced for the sake of the method rather than for the sake of the question. This is not only an esthetic disadvantage, but has very concrete mathematical consequences that are undesirable.

Examples to illustrate this remark will be clearer if compared with one of the oldest problems of the calculus of variations as a background. We mean a question that does not present any difficulties on the basis of Riemann or Lebesgue integration, but would do so for more special kinds of integrals, viz., the problem of finding the curve $x(t)$, $y(t)$, $(a \leq t \leq b)$, which joins two given points $p_0 = (x_0, y_0)$ and $p_1 = (x_1, y_1)$ and which, rotated around the x-axis, generates a surface of revolution of minimum area. Analytically speaking, we are looking for a pair of functions $x(t)$, $y(t)$, $(a \leq t \leq b)$, satisfying the conditions $x(a) = x_0$, $y(a) = y_0$, $x(b) = x_1$, $y(b) = y_1$ and minimizing the integral

$$\int_a^b y(t)\sqrt{[x'(t)]^2 + [y'(t)]^2}\,dt.$$

If the points p_0 and p_1 are sufficiently far apart, and sufficiently close to the x-axis, then the minimizing curve discovered by Goldschmidt is a broken line with two corners on the x-axis. This case is, of course, taken care of by introducing Lebesgue integration and it can be handled even on the basis of Riemann integration since the set of points at which the integrand is not defined consists of two points only. But if at the start we restricted ourselves to the consideration of functions $x(t)$ and $y(t)$ that are everywhere differentiable or, geometrically speaking, of curves with a tangent at each point, then we should exclude just that curve in which we are most interested, viz., the one that minimizes the integral. If we based the theory on a concept of integration applicable only to functions which are everywhere differentiable, the effect would be the same.

As a first example to illustrate our criticism, we consider the problem of finding two functions $x(t)$, $y(t)$ $(a \leq t \leq b)$ satisfying the conditions $x(a) = 0$, $y(a) = 0$, $x(b) = 1$, $y(b) = 0$ and minimizing the integral

$$\int_a^b \sqrt{(x^2+y^2)^3(x'^2+y'^2)+2(x^2+y^2)(xy'-yx')(xx'+yy')+(xx'+yy')^2}\,dt$$

For any curve of finite length joining the point $(0, 0)$ with the point (1, 0) the integral can be shown to assume a positive value. However, for some spirals of infinite length joining the two points the (improper) integral assumes the value zero. The existence of problems in which the minimizing curves are of infinite length was discovered by Hahn. The simple integrand mentioned above, a square root of a polynomial of degree eight, is due to Carathéodory.[8] The examples of this type show that by restricting ourselves to the consideration of curves of finite lengths at the start, we sometimes exclude just those curves in which we are mainly interested, viz., the curves minimizing the integral. On the other hand, the classical theory, whether based on Riemann or on Lebesgue integration, had to restrict itself to curves of finite lengths, or still more special curves. A method enabling us to define line integrals along all continuous curves, and to prove semicontinuity properties in such an extended domain would thus mark a distinct progress.

As a second example[9] we remember that in studying the question when $\int udx+vdy$ is independent of the path the classical theory restricts itself to functions u and v which admit partial derivatives, at least almost everywhere. Only in this case the classical condition of Cauchy, $\frac{\partial u}{\partial y}=\frac{\partial v}{\partial x}$, can be formulated. Now let $f(z)$ be a continuous function of one real variable z and set $u(x, y)=f(x+y)$ and $v(x, y)=f(x+y)$. Then the $\int udx+vdy$ is equal to $\int f(x+y)(dx+dy)=\int f(x+y)\, d(x+y)$. This integral is certainly independent of the path. For if $F(z)$ is a function satisfying the condition $F'(z)=f(z)$, then the value of the integral along a curve is equal to the difference of the values of F at the two end points of the curve. Now assume that the continuous function $f(z)$ is nowhere differentiable. Then the

$\int udx + vdy$ will be independent of the path, although neither u nor v has any partial derivative.

6. WEIERSTRASS SUMS AND GEOMETRY OF DISTANCE

These and other examples suggested another way out of the difficulties, a way parallel to the second one mentioned in the special case of the length. We shall consider the integral neither in Riemann's nor in Lebesgue's sense, but as the limit of Weierstrass sums. By a Weierstrass sum of the integral $\int_a^b F(x(t), y(t), x'(t), y'(t))dt$ associated with the set T we mean

Weierstrass sum $W(T) =$

$$\sum_{i=0}^{n-1} F\left(x(t_i), y(t_i), \frac{x(t_{i+1}) - x(t_i)}{t_{i+1} - t_i}, \frac{y(t_{i+1}) - y(t_i)}{t_{i+1} - t_i}\right)(t_{i+1} - t_i).$$

Weierstrass considered the integral as the limit to which these sums converge when $\nu(T) \to 0$. Osgood continued the work for the sake of eliminating superfluous differentiability conditions concerning the admissible curves. Bolza devotes a paragraph of his book to this idea. Tonelli studied the Weierstrass sums in a paper. But no one seems to have been interested in the general conditions concerning both the integrand and the admissible curves, to say nothing of the functional of comparison and the space, under which such a limit exists and is semicontinuous. The subsequent development of the calculus of variations moved almost entirely along the lines of Lebesgue integration. Weierstrass sums were not thoroughly studied until a few years ago, when Bouligand[10] suggested a definition of the integral as a limit of these sums, and independently the writer of this paper in a series of publications[11] developed an extensive theory of these limits applying methods developed in his former papers on metric geometry, or geometry of distance.[12]

The connection between these questions and the geometry of distance is established by a step corresponding to that from Weierstrass sums to Archimedes sums in the case of the length. Denoting by p_i the point with the coordinates $x_i = x(t_i)$, $y_i = y(t_i)$, and by $\theta_{i,\ i+1}$ the angle which the vector from p_i to p_{i+1} includes with the positive x-axis, we can write the Weierstrass sum for the integrand F associated with the finite set T in the following way:

$$\sum_{i=1}^{n-1} F(x_i,\ y_i,\ \cos\theta_{i,\ i+1},\ \sin\theta_{i,\ i+1}) \times \text{Eucl. dist. } p_i p_{i+1}$$

where $\times$ is a multiplication sign, and Eucl. dist. $p_i p_{i+1}$ denotes the Euclidean distance between p_i and p_{i+1}.

The expression can be further transformed if for any two points $p = (x, y)$ and $q = (x', y')$ we briefly write $F(p, \theta_{pq})$ instead of $F(x, y, \cos\theta_{pq}, \sin\theta_{pq})$ where θ_{pq} denotes the angle between the vector from p to q and the positive x-axis, and furthermore set $F(p, \theta_{pq}) = 0$, if $p = q$. In this notation we associate with any two points p and q of the Euclidean plane a new distance which we shall call the F-distance between p and q, viz., the number

$$F\text{-dist. } pq = F(p,\ \theta_{pq}) \times \text{Eucl. dist. } pq.$$

If an individual problem of the calculus of variations determined by an integrand $F(p, \theta)$ is under consideration, then we also shall call the F-distance the *variational distance*. In this notation the Weierstrass sum can be written in the following form entirely analogous to an Archimedes sum

$$\sum_{i=1}^{n-1} F\text{-dist. } p_i p_{i+1} \quad \text{or} \quad \sum_{i=1}^{n-1} \text{var. dist. } p_i p_{i+1}.$$

In other words, each Weierstrass sum is the length of a polygon inscribed into the curve C if the determination of the length is based on what we called the F-distance or the variational distance of pairs of points. The functional that the writer's theory associates with the curve C is the limit which the variational F-lengths of the polygons inscribed in C approach when the norms of the polygons approach zero—if there is such a limit. Our functional thus may be considered as the *F-length* or *variational length of the continuous curves*, for our derivation of the functional from the variational F-distance is exactly the same as the derivation of the Euclidean length of a curve from the Euclidean distance.

If $F(p, \theta) = 1$ for each point p and each direction θ, then F-dist. pq = Eucl. dist. pq for any two points p, q, and the variational F-length of each curve is its Euclidean length.

There is, however, one important difference between the Euclidean length and the general F-length of curves: Each curve of a Euclidean space has a finite or infinite Euclidean length, that is to say, for each curve the Euclidean lengths of the inscribed polygons converge toward a finite number or toward $+\infty$ when the norms of the polygons tend toward 0. But for certain integrands F there are curves for which the F-lengths of the inscribed polygons do not converge at all (neither toward a finite nor toward an infinite number) when the norms tend toward 0.

As an illustration of this fact, F. Alt gave the following example[13] of a continuous curve C and a continuous quasi-regular integrand F. The curve C is contained in the x-axis, and given by the equation $x(t) = t^2 \cos 2\pi/t^6$, $y(t) = 0$, $(-\sqrt[3]{2} \leq t \leq 0)$. The integrand is $F(x, \theta) = -\cos\theta \cdot x^2 \cos 2\pi/x^3$. Among the polygons of arbitrarily small norms inscribed into C there are some whose F-lengths are arbitrarily small, and some whose F-lengths are arbitrarily large. Hence, C has no F-length. Owing to its greater complexity the theory of the F-length presents difficulties which do not arise in the special case of the Euclidean length.

In the following five sections we shall deal with the application of metric geometry to the three problems about line integrals which are of interest for the calculus of variations: When does the integral exist? When is it semicontinuous? When do there exist minimizing curves? We shall see that the metric methods result in generalizations of all the four basic concepts: the integrand, the admissible curves, the functional of comparison, the space. After that we shall apply our method to the question of when the integral is independent of the path.

7. METRIC METHODS AND THE INTEGRAND

Like Tonelli and his successors we shall need continuity and regularity assumptions concerning $F(p, \theta)$. But we shall not even need lower semicontinuity of F at each point: We admit that each curve of finite length may contain a set of measure 0 consisting of points p at which $F(p, \theta)$ is not even lower semicontinuous provided that each of these exceptional points is contained in a neighborhood such that $J(C) \geq 0$ for each closed curve C contained in this neighborhood. Here, following Carathéodory, a set is said to be of measure 0 if for each $\epsilon > 0$ it can be covered by a denumerable sequence of spheres for which the sum of the diameters is $< \epsilon$. A set which can be covered by a finite set of spheres for which the sum of the diameters is arbitrarily small, will be said to be of content 0.

Moreover, we shall not need quasi-regularity of F at each point, i.e., convexity of the figurative of F for each point (x_0, y_0). For concave surfaces we can define a degree of concavity and then admit that each curve of finite length for each $\tau > 0$ may contain a set of content 0 consisting of points for which the figurative of F has a degree of concavity $\geq \tau$ provided that the degree of concavity is uniformly bounded along the whole curve.[14]

8. METRIC METHODS AND THE ADMISSIBLE CURVES

Under the aforementioned assumptions about the integrand F, the variational F-length $J(C)$ exists for each curve of finite Euclidean length and is lower semicontinuous on each set of curves whose Euclidean lengths are uniformly bounded.

Under slightly stronger conditions we can prove that the functional $J(C)$ exists *for all continuous curves* whether of finite or of infinite Euclidean length, and is lower semicontinuous for all continuous curves. The additional assumption is that for each point p there exists a neighborhood of p and a number $\omega(p) > 0$ such that for each closed polygon contained in the neighborhood the variational length is at least the $\omega(p)$-fold of its Euclidean length.[15] If F is continuous, quasi-regular, and seminormal, then it can be proved that this additional assumption is satisfied. Moreover, our theory formulates geometric conditions guaranteeing that integrands which are neither continuous nor quasi-regular, satisfy the additional assumption.[16]

While these theorems enable us to define line integrals of certain integrands and to prove their semicontinuity for all continuous curves, they are complemented by a theorem which impairs their use for the problem of finding curves which minimize the integral. For it can be shown[17] that under the aforementioned conditions along each curve whose Euclidean length is infinite the integral of F has the value $+\infty$.

However, our theory enables us to form the integral along any continuous curve even for certain integrands that are not seminormal and do not satisfy the additional assumption mentioned above. And for such integrands, the integral along a curve of infinite Euclidean length may be finite.[18] The integrands of this type which our theory enables us to deal with form a fairly large class of functions including, in particular, the integrands discovered by Hahn and Carathéodory, which we discussed in Section 5.

9. METRIC METHODS AND THE FUNCTIONAL OF COMPARISON

The reason why Tonelli and his successors referred to the length as a functional of comparison was that they had to make use of the following theorem due to Hilbert: In a compact space for each finite number M the set of all continuous curves whose lengths are $\leq M$, is compact. That is to say, from each sequence of curves whose lengths are $\leq M$ we can extract a subsequence converging toward a continuous curve. (On account of the lower semicontinuity of the length this limit curve likewise has a length $\leq M$.) The following generalization of Hilbert's theorem has been proved by the writer.[19]

In a compact space let $L(C)$ be any lower semicontinuous functional, defined (although not necessarily finite) for all continuous curves C and having the following properties:

(1) If C is a curve for which $L(C) < \infty$, and p is a point of C, then we have $L(C') + L(C'') = L(C)$ where C' and C'' denote the initial and terminal segment into which C is divided by p. If p converges toward the terminal point of C, then $L(C')$ converges toward $L(C)$.

(2) For any two distinct points of the space, p and q, the greatest lower bound of the numbers $L(C)$ for all continuous curves C joining p and q is >0.

(3) Each point p, for each $\epsilon > 0$, can be joined to each point of a sufficiently small neighborhood by a continuous curve C for which $L(C) < \epsilon$.

Then for each finite number M the set of all continuous curves C for which $L(C) \leq M$ is compact.

By virtue of this theorem, instead of the length we may use as a functional of comparison any lower continuous functional $L(C)$ with the three aforementioned properties. In particular, we may use many line integrals. For, if $G(p, \theta)$ is any function of a point p and a direction θ, and $L(C)$ is the line integral of G along C, then the functional $L(C)$ certainly satisfies the condition (1). If along each curve of finite length the integrand G is almost everywhere >0, then the line integral also satisfies condition (2). If, roughly speaking, almost everywhere along each curve of finite length, G is lower semicontinuous and has a figurative that is convex upward, then, as we know, the line integral of G is a lower semicontinuous functional. We have only to postulate that condition (3) holds, which, in fact, is a condition concerning the space (the domain of definition) rather than the functional. It states a property of the space similar to the local connectedness studied in topology. If $L(C)$ denotes the diameter of C, then condition (3) postulates that each point can be joined to each sufficiently close neighbor point by a continuous curve of arbitrarily small diameter, and that is exactly the property described in topology by the words "the space is locally connected at each point." Hence, we shall express the condition (3) by saying that the space is L-locally connected.

In particular, we can say: If G is a function which along each curve of finite length is almost everywhere >0, lower semicontinuous and quasi-regular, and the space is L-locally connected with regard to the integral $L(C)$ of G along C, then the functional $L(C)$ may serve as a functional of comparison.

Instead of postulating the uniform boundedness of the *lengths* of all curves C for which $J(C) \leq M$, it is sufficient to assume the uniform boundedness of *the values of some line integral of comparison* $L(C)$ for all curves C for which $J(C) \leq M$. Clearly this condition is weaker than the ordinary postulate.

10. GENERAL ANALYSIS AND THE EXISTENCE OF MINIMIZING CURVES

Before starting the discussion of generalized spaces we mention some consequences concerning the existence of minimizing curves in the ordinary Euclidean spaces resulting from the theorems mentioned in the preceding sections. In deriving from the lower semicontinuity of the integral the existence of curves minimizing it, we follow Tonelli's method, that is to say, we apply a theorem of the general analysis generalizing the classical theorem of R. Baire that a lower semicontinuous function on a compact domain of definition assumes its minimum. In fact, all the procedures of the calculus of variations since Hilbert's famous paper on the problem of Dirichlet which are known as "direct *methods*" are applications of *theorems* of the general analysis.

However it seemed desirable to formulate quite explicitly this theorem which implicitly underlies Tonelli's deductions. Only in this way was it possible to remove the length from its unjustified role as the only functional of comparison, and to introduce as a functional of comparison any lower semicontinuous functional of curves for which Hilbert's theorem holds. We call the theorem of the general analysis so obtained[20] "Tonelli's Principle." It reads as follows:

Let L be a limit class, i.e., a set of elements for which certain sequences of elements are said to converge toward a limit element. (In particular, each sequence all of whose elements are equal, is said to converge toward this element; moreover it is assumed that each subsequence of a sequence converging toward an element converges toward this element.)

Let β be a function defined in L, associating a (finite or infinite) number $\beta(e)$ with each element e of L, and let α be a function defined on a subset L_a of L containing, in particular, all elements e for which $\beta(e) < \infty$.

Let K be any subset of L_a closed in the set of all e for which $\beta(e) < \infty$.

If α and β satisfy the conditions

(1) For each number $\alpha_0 < \infty$ the function β is uniformly bounded above on the set of all elements e for which $\alpha(e) \leq \alpha_0$.

(2) For each $\beta_0 < \infty$ the set of all elements e for which $\beta(e) \leq \beta_0$, is compact,

then K contains at least one element e_0 that is limit of a minimizing sequence for α on K, i.e., a sequence $e_1, e_2, \cdots$ for which the numbers $\alpha(e_1), \alpha(e_2), \cdots$ converge toward the greatest lower bound of α on K.

If in addition to (1) and (2) the functions α and β satisfy the condition

(3) For each $\beta_0 < \infty$ the function α is lower semicontinuous on the set of all elements e for which $\beta(e) \leq \beta_0$,

then each element e which is limit of a minimizing sequence for α on K actually minimizes α on K.

Tonelli and his successors apply the special case of this theorem in which the elements of L are continuous curves of a Euclidean region, $\beta(e)$ is the length of the curve e, and $\alpha(e)$ the line integral in consideration formed along the curve e.

In the same way we apply the general theorem to any line integral $\alpha(e)$ satisfying the semicontinuity condition (3), and any functional of comparison $\beta(e)$ satisfying the analogue of Hilbert's theorem about the length expressed in condition (2), provided that α and β are connected by the condition (1). In this way our theorems about the semicontinuity of integrals enable us to prove in Tonelli's way the existence of curves minimizing the integral.

II. METRIC METHODS AND THE UNDERLYING SPACE

The classical theory is restricted to curves contained in a Euclidean space, or at any rate, to curves contained in a space whose points are defined by real coordinates. The integral is the variational length or F-length derived from the variational or F-distance, where var. dist. $pq = F(p, \theta_{pq}) \times$ Eucl. dist. pq.

Our theory can be developed[21] in any metric space in the sense of Fréchet. That is to say, we merely require a set of points such that with any two points a number is associated, called their geometric distance, satisfying the following conditions:

geom. dist. $pq =$ geom. dist. qp (Symmetry)
geom. dist. $pq > 0$ if $p \neq q$ (Positivity)
geom. dist. $pp = 0$ (Normality)
geom. dist. $qr +$ geom. dist. $rs \geq$ geom. dist. qs
(Triangular inequality).

The points need not be defined by coordinates, nor need the distance be expressed in terms of coordinates.[22]

We have now to say what corresponds in our metric space to the integrand $F(p, \theta)$ depending upon a point and a direction in the Euclidean case. We assume to be given a function $F(p; q, r)$ of a point p and an ordered pair of points q, r which has the value 0 if $q = r$, and we set

var. dist. pq or F-dist. $pq = F(p; pq) \times$ geom. dist. pq.

From this point on we proceed as we did in the Euclidean case. We form the F-length of a polygon, and call F-length of a continuous curve contained in the metric space the limit which the F-lengths of the inscribed polygons approach when their norms approach 0, provided that such a limit exists.

The question is under what conditions concerning F and C the F-length of a curve C exists and is a lower semi-continuous functional. In other words, what assumptions concerning $F(p; q, r)$ defined in the metric space correspond to the continuity and quasi-regularity assumptions concerning $F(p, \theta)$ defined in the Euclidean space?

Clearly the continuity and semicontinuity properties of F with regard to p can be formulated if p is a point of any general metric space. But the metric methods enable us also to express quasi-regularity properties of an integrand $F(p; q, r)$ defined in a general metric space.

The easiest way to realize this possibility starts with a second interpretation of quasi-regularity of a function $F(p, \theta)$ in the Euclidean case. First we had considered convexity properties of the figurative. The second interpretation is given in terms of the *indicatrix* of F introduced by Carathéodory. On each ray θ issuing from the point p_0 we lay off a segment of length $1/F(p_0, \theta)$. The set of all endpoints so obtained is called the indicatrix of F at p_0. In the plane the indicatrix is a curve, in the n-space a $(n-1)$-dimensional surface intersecting each ray issuing from p_0 exactly once. (If $F(p_0, \theta)=0$ for some θ, then we say that the indicatrix intersects the ray θ at its point at infinity.) Now it can be shown that in the case of a non-negative F, quasi-regularity of F at p_0 is equivalent to convexity of the indicatrix of F at p_0. On the other hand, according to a well-known theorem of Minkowski's, an $(n-1)$-

dimensional surface intersecting each ray of a bundle exactly once, is convex if and only if a certain metric associated with the surface satisfies the triangular inequality. The distance between two points q, r according to this Minkowski metric associated with the indicatrix $F(p_0, \theta)$, is the number $F(p_0, \theta_{qr}) \times$ Eucl. dist. qr. (In particular, each point of the indicatrix of F at p_0 has the distance 1 from p_0, in this metric.) Hence quasi-regularity of F at the point p_0 is equivalent to the triangular inequality

$$F(p_0, \theta_{qr}) \text{ Eucl. dist. } qr + F(p_0, \theta_{rs}) \text{ Eucl. dist. } rs \geq F(p_0, \theta_{qs}) \text{ Eucl. dist. } qs.$$

F. Alt succeeded in proving this equivalence for any integrand $F(p, \theta)$, even for the case that $F(p_0, \theta)$ is positive for some directions θ and negative for others.[23] In this case the indicatrix has to have a property called projective convexity which for a positive function F is the same as the ordinary convexity.

Now such a triangular inequality can be formulated for a function $F(p; q, r)$ defined in any metric space. We postulate

$$F(p_0; q, r) \text{ geom. dist. } qr + F(p_0; r, s) \text{ geom. dist. } rs \geq F(p_0; q, s) \text{ geom. dist. } qs$$

for any four points p_0, q, r, s. Whereupon we can prove existence and semicontinuity of the variational length of a curve in the metric space exactly in the same way as we did in the Euclidean space.

It should be noted that the triangular inequality we had to postulate does not refer to the variational distance of points. By var. dist. qr or F-dist. qr we meant the number $F(q; q, r) \times$geom. dist. qr, whereas our assumed triangular inequality deals with the numbers $F(p_0; q, r) \times$geom. dist. qr, etc. Only if $q = p_0$ this number is the variational distance between q and r. For each point p_0 the number $F(p_0; q, r) \times$ geom. dist. qr gives rise to a metric satisfying the triangular inequality. We might call this metric "tangential" to the variational metric at the point p_0.

The variational length itself does not necessarily satisfy the triangular inequality, nor is it necessarily symmetric or positive. The only one of the four postulates of Fréchet it does satisfy is the normality condition

$$\text{var. dist. } pp = 0.$$

In addition, it has those properties which result from the semicontinuity of F, the triangular inequality for the tangential metrics and the properties of the underlying geometric distance. These assumptions are sufficient to prove the existence and semicontinuity of the integral and the existence of minimizing curves to the extent we developed the theory in the Euclidean case.[24]

There is another way of looking upon the situation. The theory of length had been developed not only for the curves contained in Euclidean spaces, but far beyond; e.g., the writer had studied the length of curves contained in general metric spaces in the sense of Fréchet.[12] All that actually was needed was to even omit some of the postulates of a general metric space, few as they are, and still develop a theory of length. A development of the theory of length under weaker and weaker conditions was bound to result in a theory general enough to comprise what we called variational length of curves. This program has been carried out in a recent publication.[25]

12. UPPER SEMICONTINUITY AND MAXIMUM PROBLEMS. CONTINUITY

We saw that the integral of a function F along a curve C, in the sense of the F-length of C, is lower semicontinuous on each set of curves of uniformly bounded lengths, if the integrand F is lower semicontinuous and positively quasi-regular. (Metric methods allow us even to admit 0-sets of exceptional points.) Positive quasi-regularity of F at the point $p_0=(x_0, y_0)$ means that the figurative of F at p_0 is convex downward, or, what is equivalent, that the indicatrix of F at p_0 is convex, or, in still other words, that the following triangular inequality holds

$$F(p_0, \theta_{qr}) \times \text{Eucl. dist. } qr + F(p_0, \theta_{rs}) \times \text{Eucl. dist. } rs \\ \geq F(p_0, \theta_{qs}) \times \text{Eucl. dist. } qs.$$

It goes without saying that the integral of F is upper semicontinuous on each set of curves of uniformly bounded lengths if the integrand F is upper semicontinuous and negatively quasi-regular. By negative quasi-regularity of F at the point p_0 we mean that the figurative of F at p_0 is convex upward, the indicatrix of F at p_0 concave, and the following counter-triangular inequality holds

$$F(p_0, \theta_{qr}) \times \text{Eucl. dist. } qr + F(p_0, \theta_{rs}) \times \text{Eucl. dist. } rs \\ \leq F(p_0, \theta_{qs}) \times \text{Eucl. dist. } qs.$$

It is further obvious that with each theorem about the existence of minimizing curves derived from the lower semicontinuity of the integral there corresponds a theorem about the existence of maximizing curves derived from the upper semicontinuity of the integral.

Continuity of the integral is equivalent to its being both lower and upper semicontinuous. Sufficient for the continuity of the integral on each set of curves of uniformly bounded lengths is thus that the integrand is both lower and upper semicontinuous, and both positively and negatively quasi-regular. This means that the integrand is continuous and its figurative at each point is flat, its indicatrix at each point is flat, and the following triangular equality holds for each p_0

$$F(p_0, \theta_{qr}) \times \text{Eucl. dist. } qr + F(p_0, \theta_{rs}) \times \text{Eucl. dist. } rs = F(p_0, \theta_{qs}) \times \text{Eucl. dist. } qs.$$

In the two-dimensional case of an integrand $F(x, y, u, v) = F(p, \theta)$ the figurative at the point (x_0, y_0), that is the surface $z = F(x_0, y_0, x', y')$, is flat (i.e., a plane through the origin of the (x', y', z)-space) if and only if for two constants u_0 and v_0 we have $F(x_0, y_0, x', y') = u_0x' + v_0y'$. Then the indicatrix has the following equation in polar coordinates: $\rho = 1/F(x_0, y_0, \cos\theta, \sin\theta) = 1/(u_0\cos\theta + v_0\sin\theta)$ which is the equation of a straight line.

Sufficient for the continuity of the integral on each set of curves of uniformly bounded lengths is thus that the integrand $F(x, y, \cos\theta, \sin\theta)$ is a continuous function of the form $u(x, y)\cos\theta + v(x, y)\sin\theta$, which is the case if $u(x, y)$ and $v(x, y)$ are continuous functions. The integral of this integrand along the curve C is usually written in the form $\int_C u(x, y)dx + v(x, y)dy$ or, briefly, $\int_C udx + vdy$.

13. LINE INTEGRALS INDEPENDENT OF THE PATH

We now turn to problem IV of section 2: Under what conditions is a line integral independent of the path, i.e., when does it assume the same value for any two coterminal curves of finite length? Such an integral clearly is a continuous functional. We saw in the last section that it is sufficient for the continuity of an integral that it be of the form $\int u(x, y)dx + v(x, y)dy$, where u and v are continuous functions. This condition is not necessary for the continuity of the integral. But although integrals of the form $\int udx + vdy$ are not the only continuous integrals, they seem to be the most interesting ones. Thus, as it is customary, we shall restrict the research to a study of the conditions under which an integral of the form $J(C)=\int_C udx+vdy$ is independent of the path C. These integrals have the further important property that $J(-C)=-J(C)$, where $-C$ denotes the curve C traversed in the opposite direction.

If the first classical condition mentioned in section 3 holds, that is to say, if $\partial u/\partial y$ and $\partial v/\partial x$ exist, are continuous, and are equal for each (x, y) of a closed and bounded domain D in the (x, y)-plane, then the functions $\partial u/\partial y$ and $\partial v/\partial x$ are uniformly continuous, and for each $\epsilon>0$, one readily proves: There is a number $l_\epsilon>0$, such that if

$$(\underline{a}, \underline{b}),\ (\underline{a}, \bar{b}),\ (\bar{a}, \underline{b}),\ (\bar{a}, \bar{b})$$

are the vertices of any rectangle whose diagonal is $<l_\epsilon$, then the following inequality, to which we shall refer as the *fundamental ϵ-inequality*, holds

$$\left|\frac{u(\underline{a}, \bar{b})-u(\underline{a}, \underline{b})}{\bar{b}-\underline{b}}-\frac{v(\bar{a}, \underline{b})-v(\underline{a}, \underline{b})}{\bar{a}-\underline{a}}\right|<\epsilon$$

or, what is equivalent,

$$|[v(\underline{a},\underline{b})(\bar{b}-\underline{b})+u(\underline{a},\bar{b})(\bar{a}-\underline{a})]-[u(\underline{a},\underline{b})(\bar{a}-\underline{a})+v(\bar{a},\underline{b})(\bar{b}-\underline{b})]|< \epsilon(\bar{a}-\underline{a})(\bar{b}-\underline{b}).$$

Now let R be the rectangle $\underline{x}\leq x\leq\bar{x}$, $\underline{y}\leq y\leq\bar{y}$. By a rectangular net in R, we mean a matrix of points (x_i, y_j) $(i=0, 1, \cdots, m;\ j=0, 1, \cdots, n)$ for which

$$x_0=\underline{x},\ x_m=\bar{x},\ x_i<x_{i+1};\ y_0=\underline{y},\ y_n=\bar{y},\ y_j<y_{j+1}.$$

Each rectangle

$$(x_i, y_j),\ (x_{i+1}, y_j),\ (x_i, y_{j+1}),\ (x_{i+1}, y_{j+1})$$

is called a mesh of the aforementioned net, N. The largest diagonal of a mesh, i.e., the largest of the numbers $\sqrt{(x_{i+1}-x_i)^2+(y_{j+1}-y_j)^2}$, will be called the norm of the net N.

Now let us denote by R_1 and R_2 the two curves contained in the boundary of R, starting at the point $(\underline{x}, \underline{y})$, and terminating at the point $(\bar{x}, \bar{y})$. Let R_1 consist of the left and upper, R_2 of the lower and the right side of R. We shall set

$$J(R_1) = \int_{R_1} udx + vdy \text{ and } J(R_2) = \int_{R_2} udx + vdy.$$

Let us furthermore denote by $T_1(N)$ and $T_2(N)$ the ordered sets of points of the net N on R_1 and R_2.

$$T_1(N): (\underline{x}, y_0), (\underline{x}, y_1), \cdots, (\underline{x}, y_{n-1}), (\underline{x}, \bar{y}), (x_1, \bar{y}), \cdots, (x_m, \bar{y})$$
$$T_2(N): (x_0, \underline{y}), (x_1, \underline{y}), \cdots, (x_{m-1}, \underline{y}), (\bar{x}, \underline{y}), (\bar{x}, y_1), \cdots, (\bar{x}, y_n)$$

Obviously, the norms of $T_1(N)$ and $T_2(N)$ are smaller than the norm of the net N.

Now let us assume that for some $\epsilon > 0$, the fundamental ϵ-inequality holds for each sufficiently small rectangle contained in the rectangle R. Then it is easy to see that the Weierstrass sums for the integrals $J(R_1)$ and $J(R_2)$, associated with the sets $T_1(N)$ and $T_2(N)$, differ by less than $2\epsilon(\bar{x} - \underline{x})(\bar{y} - \underline{y})$. Consequently, if for each $\epsilon > 0$ the fundamental ϵ-inequality holds for each rectangle contained in R, then the Weierstrass sums of $J(R_1)$ and $J(R_2)$ associated with sets of arbitrarily small numbers differ by as little as we please, and, hence, $J(R_1) = J(R_2)$, from which it easily follows that, within the rectangle R, the integral $J = \int udx + vdy$ is independent of the path.

An analysis of the outlined proof shows[26] that the crucial equality $J(R_1) = J(R_2)$ can be derived from a weaker assumption than the condition that the fundamental ϵ-inequality holds for each sufficiently small rectangle contained in R. It is entirely sufficient to know that for each ϵ there is one rectangular net in R whose norm is less than ϵ, and each of whose meshes satisfies the fundamental ϵ-inequality. This condition is much weaker than the first classical assumption, from which it can be derived. It holds even in many cases in which neither u nor v have any partial derivatives whatever. For instance, it holds in the case of the integral $\int f(x+y)\ (dx+dy)$, where f is a continuous, nowhere differentiable function of one real variable—the case discussed in section 5, which cannot be covered by the classical theory since the integrand does not admit partial derivatives. For this integral, the fundamental ϵ-inequality is satisfied with regard to each square whose sides are parallel to the axes. If

$$(a,\ b),\quad (a,\ b+s),\quad (a+s,\ b),\quad (a+s,\ b+s)$$

are the corners of such a square, then the left side of the fundamental ϵ-inequality for this square is

$$\left|\ [f(a+b)s+f(a+b+s)s]-[f(a+b)s+f(a+s+b)s]\ \right|.$$

Since this expression is obviously equal to zero, we see that for each square whose sides are parallel to the axis, the fundamental ϵ-inequality holds even for each ϵ. Consequently, the fundamental ϵ-inequality holds for each mesh of a net all of whose meshes are squares, and since there are such quadratic nets of arbitrarily small norms, it is clear that our sufficient condition holds, which guarantees that the integral is independent of the path.

The sufficient condition for the independence of the integral that we mentioned above is, however, not necessary. If we consider the following integral

$$\int|\ (x-y)\ |\ (dx-dy),$$

then it is clear that the integral is independent of the path and yet no rectangle whose left lower corner is the origin satisfies the fundamental ϵ-inequality for any number $\epsilon < 2$. Consequently, if we denote by R the unity square in the first quadrant, then there is no rectangular net in R all of whose meshes satisfy the fundamental ϵ-inequality for sufficiently small values of ϵ.

However by considering what we call dotted nets,[26] instead of nets, we can formulate a condition that is both sufficient and necessary in order that the integral $\int udx + vdy$ be independent of the path. By a dotted net we mean a rectangular net on the boundaries of whose meshes finite ordered sets of points are inserted. Let

$$(\underline{a}, \underline{b}),\ (\bar{a}, \underline{b}),\ (\underline{a}, \bar{b}),\ (\bar{a}, \bar{b})$$

be the vertices of a rectangle M and let

$$a_0 = \underline{a} < a_1 < \cdots < a_{m-1} < a_m = \bar{a};$$
$$b_0 = \underline{b} < b_1 < \cdots < b_{n-1} < b_n = \bar{b}.$$

By the dotted rectangle M, we mean the four vertices of M together with the ordered set of points $(\underline{a}, b_1)$, $(\underline{a}, b_2)$, $\cdots$, $(\underline{a}, b_{n-1})$ on the left side, and three corresponding ordered sets on the other three sides of M. We shall say that this dotted rectangle M satisfies the strong ϵ-inequality if

$$\Big|\Big[\sum_{j=0}^{n-1} v(\underline{a}, b_j)(b_{j+1} - b_j) + \sum_{i=0}^{m-1} u(a_i, \bar{b})(a_{i+1} - a_i)\Big] -$$
$$\Big[\sum_{i=0}^{m-1} u(a_i, \underline{b})(a_{i+1} - a_i) + \sum_{j=0}^{n-1} v(\bar{a}, b_j)(b_{j+1} - b_j)\Big]\Big|$$
$$< \epsilon(\bar{a} - \underline{a})(\bar{b} - \underline{b}).$$

In order that the integral $\int udx + vdy$ be independent of the path, it is necessary and sufficient that for each ϵ, there exist a dotted net whose norm is less than ϵ and each of whose dotted meshes satisfies the strong ϵ-inequality. In a way, our necessary and sufficient condition seems to be not so different from the statement that the integral is independent of the path, but this is only apparent. In order to understand the great difference between our condition and the independence statement, one has to realize that our condition refers for each ϵ to only one net whose norm is less than ϵ and which satisfies the strong ϵ-inequality. How weak this condition is can be seen from the fact that it is much weaker than the sufficient condition concerning undotted nets that we had studied before, which in turn was much weaker than the first classical assumption, since it took care of cases in which the integrand does not have any partial derivatives. By considering one sequence of dotted nets, the k-th of which has a norm less than $1/k$ and satisfies the strong $1/k$-inequality, we reach a condition still sufficient, but so weak that it is also necessary. It should be noticed that this condition can be directly inferred[26] also from the second classical condition mentioned in section 3, the Goursat-Heffter condition dealing with functions u and v that have differentials, although it is hardly necessary to remark that on the other hand our condition is much weaker than also the second classical assumption. In a recent paper Fubini[27] formulated a necessary and sufficient condition of the type of the one presented above but dealing with a simpler concept of dotted nets.

What has this theory to do with metric geometry? It is clear from what we said that our conditions can be formulated without referring to metric concepts. But it is also clear that the introduction of a metric terminology is helpful and suggestive, and that the leading ideas of our theory are suggested by the methods of metric geometry.

The fundamental ϵ-inequality for a rectangle simply means that the variational polygonal length of the polygon R_1, and the variational polygonal length of the polygon R_2 differ by a quantity that is small compared with the area of R. Here R_1 is the polygon consisting of the left lower, the left upper, and the right upper vertices of R; and R_2 is the polygon consisting of the left lower, the right lower and the right upper vertices of R. The strong ϵ-inequality for a dotted rectangle states that the variational polygonal lengths of two polygons containing R_1 and R_2, respectively, have a difference which is small compared with the area. It is interesting and instructive to realize the geometric meaning of each step of the proofs which, in fact, were found by such geometric considerations.

14. RÉSUMÉ

Owing to the limitations of a lecture, our exposition of the metric methods in analysis must necessarily be incomplete in many ways.

More results can be obtained along these geometric lines than could be mentioned here. It is clear that what we have said about integrals of real functions independent of the path has a bearing on the theory of functions of one or more complex variables. Furthermore, metric methods yield new proofs for the classical theorems of vector analysis, in most cases under conditions weaker than the classical ones. Finally, our methods can be applied to multiple integrals.

All we could cover in this lecture was the enumeration of some of the results, restating classical theorems under weaker conditions. Taken separately, some of these generalizations may seem less significant. What one should have in mind is that our metric methods lead to progress concerning line integrals in many directions at the same time. (Naturally we did not even mention progress due to these methods in other fields of analysis, as, for instance, in differential geometry, of which a résumé is to be found in L. M. Blumenthal's book, *Distance Geometries*.[12])

Another aspect which may be as important as the generalizations of the classical results is the modification of their proofs due to the introduction of metric methods, and this point naturally had to be omitted in a single lecture altogether; but an analyst, studying the metric proofs of the generalized classical results, will notice considerable deviations from the classical proofs.

One of the characteristics of metric proofs is that one can visualize each step of the argument. What these methods try to introduce into analysis is a complete *geometrization*. With regard to the existence theorems of the calculus of variations, this aim has been accomplished. It is a well known fact that some analysts are not too fond of geometric methods. (Twenty years ago, an eminent analyst was reported to have defined geometry as that branch of mathematics in which false statements are called evident.) If one does not like geometric intuition accompanying analytic arguments, then one need not evoke the visualization of the metric arguments in analysis, but if one does like it one may.

Another characteristic of metric proofs is the *logicalization* of the theory which is due to them. They try to derive statements as comprehensive as possible from a minimum of assumptions, in the simplest way. Only such hypotheses are assumed as are actually used, and the proofs make it clear where and to what extent the assumptions are used. Metric solutions of problems of analysis are of the type of those postulational theories in which the arguments are interlocked like the wheels of a watch of greatest efficiency and simplicity. In particular, they do not follow the century-old use of starting with differentiability assumptions—of starting with them in order to be able to apply the calculus. (The legend does not say whether a geometrist retorted to the analyst's definition of geometry, that analysis is that branch of mathematics which derives interesting conclusions from superfluous assumptions!)

It is one of the wonderful achievements of calculus that for many individual curves which are given by simple functions, we can find the length by simple processes of differentiation and integration, the differentiation applicable only to curves with tangents. E.g., we can say right away that the arc of the parabola given by the equation $y=x^2/2$ $(0 \leq x \leq a)$ has the length

$$\int_0^a \sqrt{1+x^2}\,dx = \frac{a}{2}\sqrt{1+a^2} + \frac{1}{2}\log(a+\sqrt{1+a^2}).$$

Statements of this type which are of the highest importance will always remain the domain of calculus. But for certain purposes, we are also interested in statements concerning length which are of another type, namely general statements about properties of the length of all curves, e.g., that the length is lower semicontinuous, that is to say, that all curves sufficiently close to a curve C of finite length can not be much shorter than C. Why should we, in proving a theorem of this type which holds for *all* curves, restrict ourselves to the consideration of curves whose lengths can be computed by certain simple processes, important as this computation may be for those curves to which it is applicable? Why should we artificially base the study on assumptions necessary for the application of calculus but without any bearing on the problem? The enormous successes of calculus in its proper domain should not induce us to apply its ideas beyond this proper domain. While computations of quantitative properties of many simple individual objects belong to the domain, many comprehensive statements about all objects of more general type seem to lie beyond the scope of classical analysis. Classical analysis frequently can obtain general statements of great heuristic value, but in some cases not the definite answers.

The metric attack on such problems is a very direct one, admitting only the restrictions necessitated by the problem, without introducing limitations which are due to the method.

NOTES

[1]Short proofs of these theorems are to be found, e.g., in Tonelli's *Fondamenti di Calcolo delle Variazioni*, Bologna, 1921, and in Saks's *The Theory of the Integral*, New York, 1937.

[2]The most thorough investigations of these questions seem to be contained in the papers by Pringsheim in the *Sitzungsber. Akad. Wiss., München*, (1895–1903).

[3]*Math. Ann.*, LXIII (1907), p. 254.

[4]*Duke Math. Jour.*, Vol. 4 (1938), p. 132.

[5]See *Ergebnisse e. math. Kolloquiums*, Vienna, 8, 1937, p. 17 (footnote).

[6]*Ann. Ec. Norm.*, t. 24 (1907).

[7]*Acta Litt. Sci.*, Szeged, IV (1929), p. 38.

[8]*Variationsrechnung*, Leipzig, 1935, p. 310.

[9]See the author's paper in the *Proc. Nat. Acad. Sci.*, Vol. 25 (1939), p. 621.

[10]*Essai sur l'unité des méthodes directes*, Bruxelles, Hayez, 1933, p. 32.

[11]*Fundamenta Math.*, XXV (1935), p. 441; *Comptes Rendus*, Paris, t. 201 (1935), p. 705; t. 202 (1936), p. 1007, p. 1648; *Proc. Nat. Acad. Sci.*, Vol. 23 (1937), p. 244; Vol. 25 (1939), p. 621; and the comprehensive paper, "Metrische Geometrie und Variationsrechnung," *Ergebnisse e. math. Kolloquiums*, 8, 1937, p. 1.

[12]In particular, the paper in the *Math. Ann.*, CIII (1930), pp. 466–501. A survey of the main results of metric geometry and its applications up to the year 1938 is contained in L. M. Blumenthal's book, *Distance Geometries: A Study of the Development of Abstract Metrics*, The University of Missouri Studies, Vol. 13 (1938). See also the author's paper, "La géométrie des distances et ses relations avec les autres branches des mathématiques," *L'Enseignement Math.*, t. 35 (1936), p. 348. Interesting mathematical and historical remarks are to be found in a paper by Pauc in the *Ann. Sc. Norm. Sup.*, Pisa, VIII (1939), concerning what Pauc calls the Weierstrass-Bouligand-Menger integral and its relations to the Riemann and Lebesgue integrals.

[13]*Ergebnisse e. math. Kolloquiums*, 8, 1937, p. 34.

[14]See the definition of $\tau\ (p,\ \rho)$, *loc. cit.*, p. 11 (§11), and the assumptions III and IV, p. 14 and p. 28.

[15]*Loc. cit.*, p. 19 (§ 28) and the theorems on p. 20.

[16]See the condition V* on p. 29 of the paper, *loc. cit.*

[17]*Loc. cit.*, p. 20 (§ 31).

[18]*Loc. cit.*, Chapter VI, pp. 29–32.
[19]*Loc. cit.*, pp. 21–23.
[20]*Loc. cit.*, Chapter II, p. 13, and *Proc. Nat. Acad. Sci.*, Vol. 25 (1939), p. 621.
[21]*Loc. cit.*, and *Proc. Nat. Acad. Sci.*, Vol. 23 (1937), p. 244.
[22]The results may be applied to Hilbert space. See Wald *loc. cit.* note 13, p. 36.
[23]*Loc. cit.*, note 13, p. 32.
[24]*Proc. Nat. Acad. Sci.*, Vol. 23 (1937), p. 244.
[25]*Loc. cit.*, Vol. 25 (1939), p. 621.
[26]*Proc. Nat. Acad. Sci.*, Vol. 25 (1939), p. 621.
[27]*Proc. Nat. Acad. Sci.*, Vol. 26 (1940), p. 199.